Detecting and Segmenting White Blood Cells in Microscopy Images of Thin Blood Smears

Golnaz Moallem*, Mahdieh Poostchi†, Hang Yu†, Kamolrat Silamut‡, Nila Palaniappan§
Sameer Antani†, Md Amir Hossain¶, Richard J Maude‡, Stefan Jaeger†, and George Thoma†

*Electrical Engineering Department, Texas Tech University
Lubbock, TX 79409, USA (e-mail: golnaz.moallem@ttu.edu)

†Lister Hill National Center for Biomedical Communications
U.S. National Library of Medicine, Bethesda, MD 20894, USA
(e-mail: {mahdieh.poostchi, hang.yu, sameer.antani, stefan.jaeger, george.thoma}@nih.gov)

‡Mahidol-Oxford Tropical Medicine Research Unit
Bangkok 10400, Thailand (e-mail: richard@tropmedres.ac, ksilamut@gmail.com)

§University of Missouri-Kansas City
Kansas City, MO 64110, USA (e-mail: nila.palaniappan@gmail.com)

¶Chittagong Medical College & Hospital
Chittagong, Bangladesh

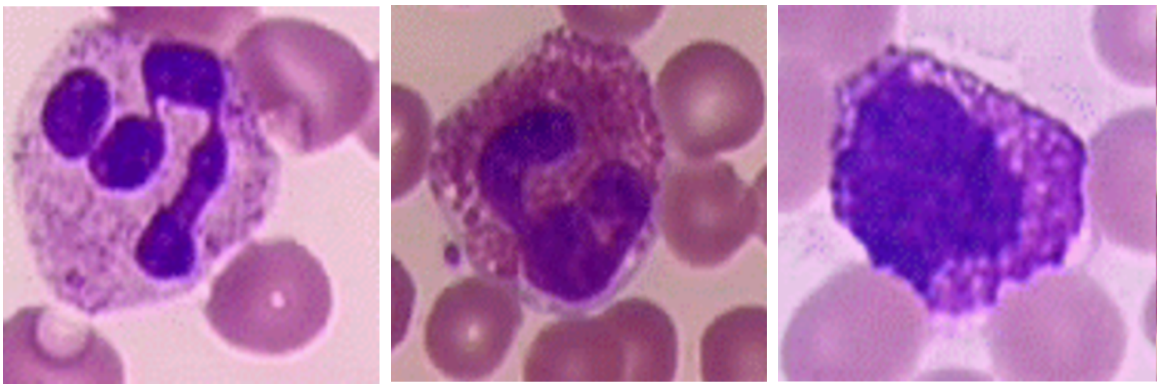
***Abstract*—A malarial infection is diagnosed and monitored by screening microscope images of blood smears for parasite-infected red blood cells. Millions of blood slides are manually screened for parasites every year, which is a tedious and error-prone process, and which largely depends on the expertise of the microscopists. We have developed a software to perform this task on a smartphone, using machine learning and image analysis methods for counting infected red blood cells automatically. The method we implemented first needs to detect and segment red blood cells. However, the presence of white blood cells (WBCs) contaminates the red blood cell detection and segmentation process because WBCs can be miscounted as red blood cells by automatic cell detection methods. As a result, a preprocessing step for WBC elimination is essential. Our paper proposes a novel method for white blood cell segmentation in microscopic images of blood smears. First, a range filtering algorithm is used to specify the location of white blood cells in the image following a Chan-Vese level-set algorithm to estimate the boundaries of each white blood cell present in the image. The proposed segmentation algorithm is systematically tested on a database of more than 1300 thin blood smear images exhibiting approximately 1350 WBCs. We evaluate the performance of the proposed method for the two WBC detection and WBC segmentation steps by comparing the annotations provided by a human expert with the results produced by the proposed algorithm. Our detection technique achieves a 96.37% overall precision, 98.37% recall, and 97.36% F1-score. The proposed segmentation method grants an overall 82.28% Jaccard Similarity Index. These results demonstrate that our approach allows us to filter out WBCs, which significantly improves the precision of the cell counts for malaria diagnosis.**

I. Introduction

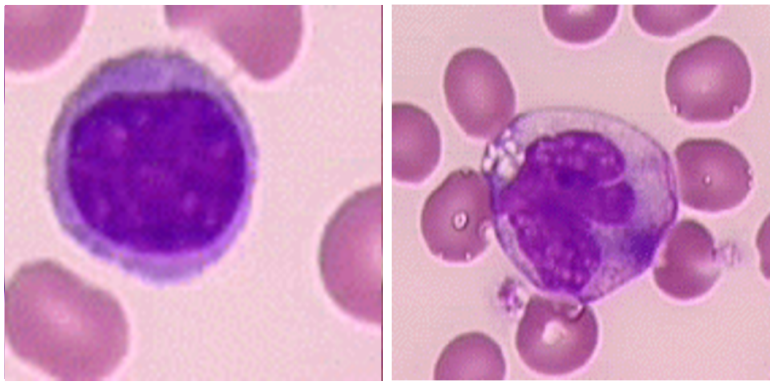
Microscopy is the main technique for malaria diagnosis. Millions of blood slides are screened using microscopy every year for diagnosing and monitoring malaria disease. Analyzing the huge volume of slides manually is extremely labor intensive and subjective to some extent. Therefore, developing a software for diagnosis of malaria greatly assists with supplementing or even substituting the manual process. The software that we developed runs on a standard Android smartphone that is attached to a microscope by a low-cost adapter. Utilizing the smartphone's built-in camera, this software captures the images of thin blood smear slides through the eyepiece of the microscope. The smartphone application performs an automatic screening of malaria parasites. The first step of the method implemented in this application is to detect and segment red blood cells. However, the presence of white blood cells (WBCs) is adversely affecting the precision of the red blood cell detection and segmentation procedure since WBCs are often mistaken for red blood cells by the implemented algorithm. Consequently, a preprocessing step for WBC elimination is necessary. White blood cells, also known as "Leukocytes" are nucleated cells in the blood that protect the body against infection diseases. There are five main types of WBCs: Neutrophils, Eosinophils, Basophils, Lymphocytes, Monocytes. Fig. 1 shows the typical appearance of these five categories in images of thin blood smears. In microscopic images of thin blood smears, morphological properties of the nucleus, the texture of the cytoplasm, and the size and shape of different WBC types varies. Besides, there may be an uneven illumination pattern introduced to the image. The staining process may also result in a diverse range of staining shades and so, the WBCs demonstrate different staining shades in our set of blood smear images. Therefore, segmenting WBCs in images of thin blood smears can be a complex and challenging task.

Detection and segmentation of white blood cells in microscopic images of thin blood smears has been widely studied. Many methods employing active contours [1], [2],

978-1-5386-1235-4/17/$31.00



(a) Neutrophils (b) Eosinophils (c) Basophils



(d) Lymphocytes (e) Monocytes

Fig. 1. Five main types of white blood cells: (a) Neutrophils, (b) Eosinophils, (c) Basophils, (d) Lymphocytes and (e) Monocytes (http://www.uwosh.edu/med_tech/what-is-elementary-hematology/white-blood-cells)

[3], color and feature extraction approaches [4], algorithms based on fuzzy morphological operations [5] and clustering techniques [6], [7] have been developed for the WBC segmentation purpose. In particular, Hou et al. [8] proposed an algorithm that performs WBC segmentation with BandMax and spectral angle mapping as a preprocessing step to divide the boundaries between the cells followed by an SVM (support vector machine) cell screening algorithm to segment the hyperspectral cell images. Prinyakupt et al. [9] introduced a technique for WBC segmentation comprising a preprocessing step, nucleus segmentation and cell segmentation. The segmentation algorithm combined thresholding, white blood cells' morphological properties in morphological operation and ellipse curve fitting to segment the cells present in the slide images, using the calibrated size of a real cell relative to image resolution. Manik et al. [10] proposed a framework for cell segmentation that followed an RGB-to-gray-scale conversion followed by an adaptive histogram equalization and binary conversion to acquire the binary mask of the image. Afterwards, morphological operations such as hole filling and image opening are executed to segment the WBCs in the image, employing the corresponding binary mask. Dorini et al. [11] implemented a novel method to segment the nucleus and then the cytoplasm of each WBC. To segment the nucleus, an image preprocessing step applying the self-dual multiscale morphological toggle (SMMT) for contour regularization is performed followed by a Watershed transform to estimate the boundaries of the nucleus. The algorithm follows two different schemes based on granulometric analysis and morphological transformations to estimate the cytoplasm region. Zhang et al. [12] demonstrated a nucleus and cytoplasm segmentation method for the WBCs. Their segmentation algorithm combines a color adjustment step with color space decomposition and k-means clustering for segmentation. Li et al. [13] proposed a dual-threshold method based on a strategic combination of RGB and HSV color space for WBC segmentation. In this algorithm, the contrast-stretched gray version of the image and the H component image from the transformed HSV color space are used in a dual-threshold method for determining the optimal thresholds followed by the image thresholding process. Then, the resultant thresholded image is denoised employing mathematical morphology operations and median filtering.

II. Methodology

In this work, we present a novel technique that successfully performs detection and segmentation of WBCs. We follow three main steps in this algorithm to fulfill the purpose of WBC segmentation: a preprocessing step that extracts the region of

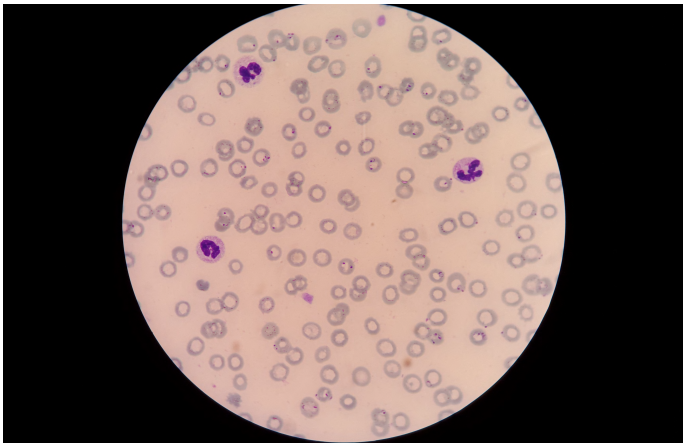


Fig. 2. A sample microscopy slide image of a thin blood smear

interest (ROI), a cell detection step using a range filtering method, and a cell segmentation step utilizing level-set active contours. In the preprocessing step, the ROI in the image is extracted. The cell detection step localizes any WBC present in the image using a range- filtered version of the image. Finally, in the cell segmentation step, the detected white blood cells are segmented utilizing a level-set method.

A. ROI Detection

In the first step, we extract the ROI of a microscopy slide image of a thin blood smear. A sample slide image is presented in Fig. 2. The region of interest is the region of the slide image that is actually visible through the eyepiece of the microscope. In other words, in this step, the unwanted black background is eliminated. First, the image is subsampled to reduce the computational complexity. Then, the image is converted to grayscale by removing the hue and saturation information while maintaining the luminance. After this, the binarized version of the image is derived by replacing all the intensity values above a globally determined threshold with ones and allocating all the other values to zeros. The global threshold for binarization is derived using Otsu's method. The purpose of binarization is to acquire the binary mask of the subsampled image, in which the ROI appears as a big white blob. Finally, the portion of the image that has nonzero values in the binarized mask is extracted and the remaining pixels are discarded. As the result, the part of the image that contains the visible cell region is extracted. Fig. 3 demonstrates the result of this step.

B. WBC Detection Using Range Filtering

In the second step, any white blood cell present in the extracted ROI from the previous step is detected using a range filtering technique. With the use of a range filter, for each pixel, the intensity value is replaced with the difference value of maximum and minimum intensities within the range of the filter. For example, for a range filter of size 3x3, the intensity value of each pixel is determined as the difference of highest and lowest intensity values in the 3x3 neighborhood of the

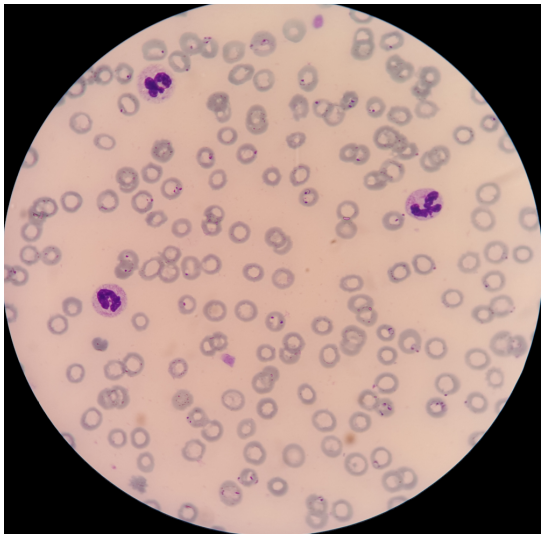


Fig. 3. The extracted ROI from the slide image in Fig. 2

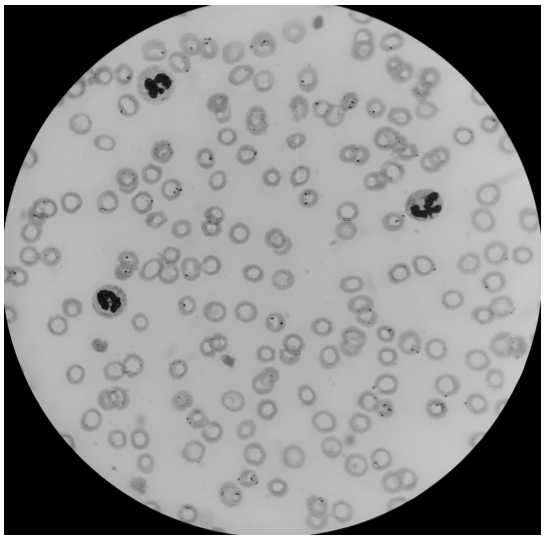


Fig. 4. The green channel of the extracted ROI in Fig. 3

pixel. Range filtering performance resembles average filtering; however, the edges are preserved in this case. Considering the green channel of the extracted ROI, shown in Fig. 4, white blood cells' nuclei present relatively higher contrast with the neighboring pixels compared to red blood cells. Therefore, at the edges of these nuclei, the range-filtered image has higher values compared to other parts of the image. In other words, the range filtering highlights a nucleus' edge pixels, and thus provides us a proper tool to distinguish the corresponding nuclei pixels from the rest of the extracted ROI. Fig. 5 presents the range-filtered version of the extracted ROI. Since nucleus edge pixels have comparatively higher intensity values, an intensity thresholding draws out the edge pixels from the rest of the image. The result is a binary mask with nucleus edge pixels having values of one. A follow-up morphological filling operation fills the empty space within the edge pixels resulting

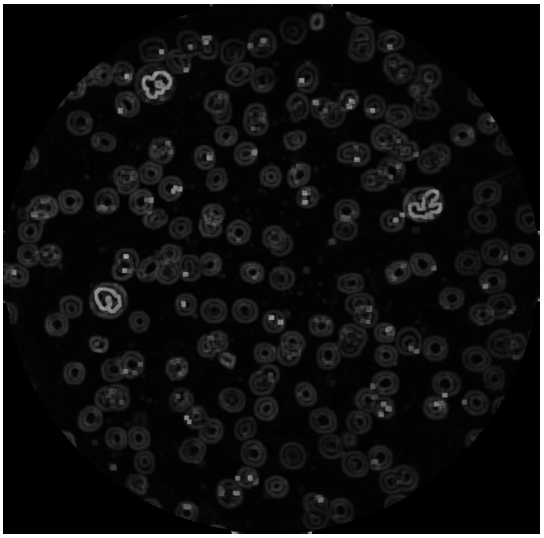


Fig. 5. The range-filtered ROI

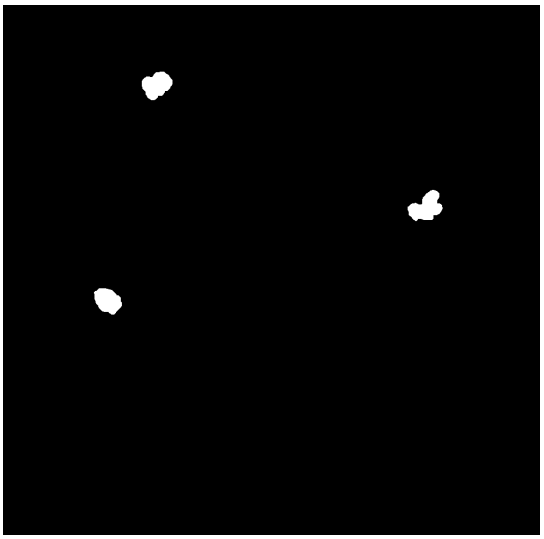


Fig. 6. The resultant binary mask with nuclei appearing as white blobs

in a binary mask, with approximations of nuclei appearing as white blobs. These approximations represent partial WBCs. Fig. 6 demonstrates the resultant binary mask.

C. WBC Segmentation Step Using Level-set Algorithm

The last step estimates the boundary pixels of the detected WBCs employing the Chan-Vese level-set algorithm. The Chan-Vese level-set, also known as "Active Contours without Edges" is a model for active contours to detect objects in an image employing techniques of curve evolution, Mumford-Shah functional for segmentation and level sets [14]. The detected partial WBCs present in the binary mask derived from the previous step are used to segment the WBCs present in the image employing the Chan-Vese algorithm. One cell at a time, a mask is generated where only the nucleus boundary pixels of the corresponding cell are white. In this mask, the initial contour is evolved outwards in the green channel of the extracted ROI. The evolution of the Chan-Vese level-set is terminated when it reaches the boundary pixels of the corresponding WBC. The resultant curve is the estimated boundary of the WBC. This process is repeated for all the detected WBCs. The estimated boundaries are marked green in the extracted ROI in Fig. 7.

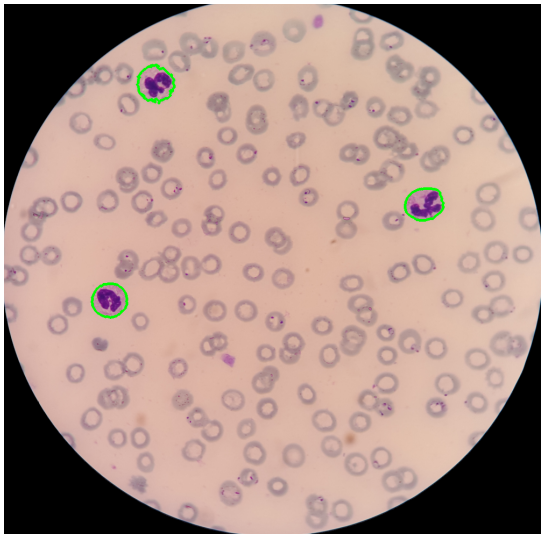


Fig. 7. The estimated WBC boundaries shown in green on the original slide image.

III. Results & Discussion

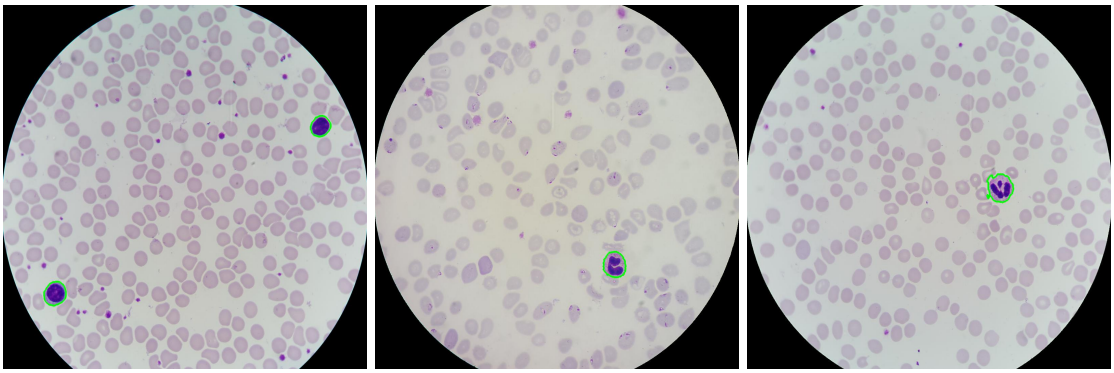
A. Image Dataset

The dataset that is used in this work contains archived thin blood smear images acquired from Chittagong Medical College & Hospital in Bangladesh [15]. This set contains more than 1300 images containing approximately 1350 WBCs. The average number of WBCs in an image is 1.12. All images are in RGB color space. The full set was acquired using a smartphone mounted on a regular light microscope. Thus, the visual region in the image is what is captured through the microscope. The visual region extraction is taken care of in the preprocessing step of the proposed algorithm. All of the WBCs in the entire dataset are annotated manually by an experienced slide reader using the Firefly annotation tool (firefly.cs.missouri.edu).

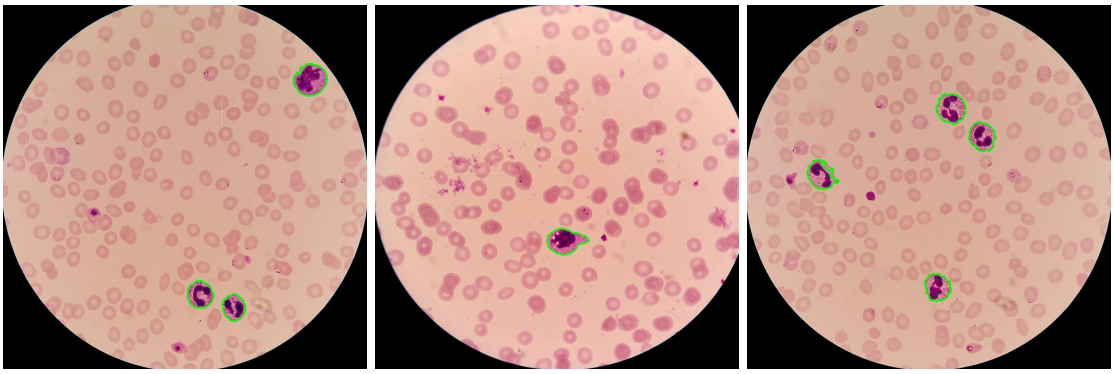
B. Qualitative Results

Fig. 8 demonstrates the results of the segmentation step with estimated WBC boundaries marked green in various image slides from the dataset.

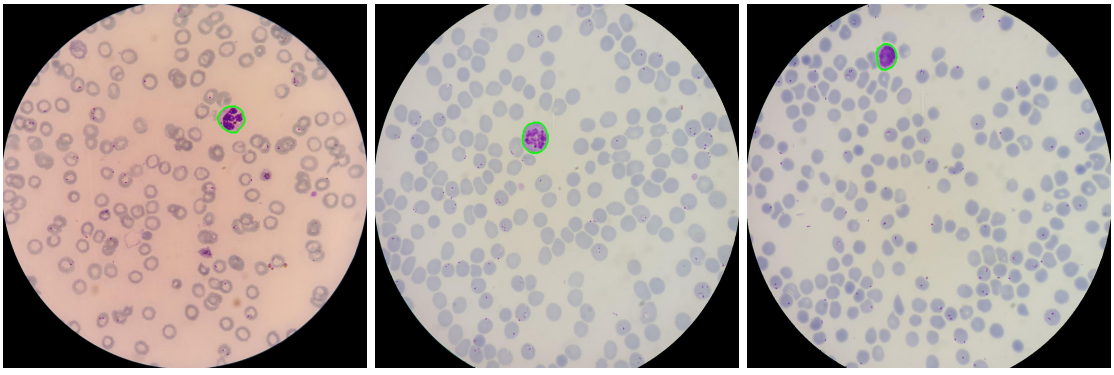
As shown, the algorithm is capable of segmenting distinct types of WBCs, with diverse nucleus morphologies and cytoplasm textures and performs robustly to various staining shades. The proposed technique can handle segmentation of multiple WBCs in a single slide image. The proposed method can also estimate WBC edge pixels even if the WBC is touching red blood cells.



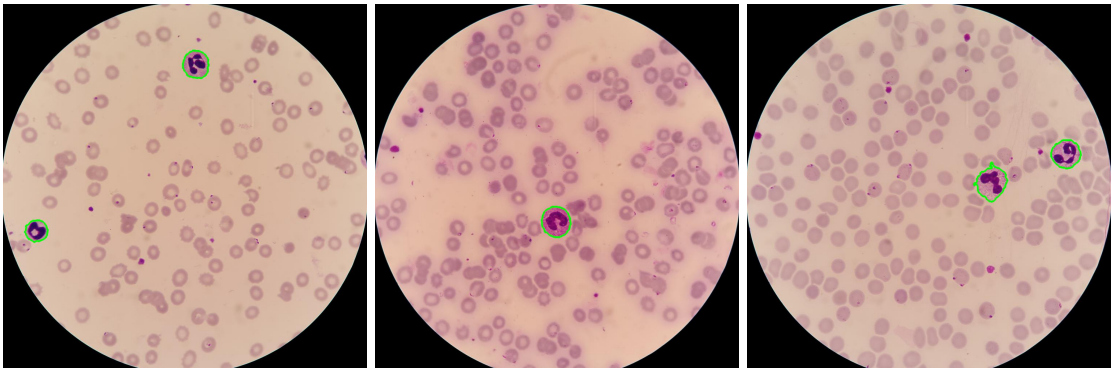
(a) (b) (c)



(d) (e) (f)



(g) (h) (i)



(j) (k) (l)

Fig. 8. The segmentation results marked on various sample slide images

These results show the strength of the proposed algorithm in segmenting WBCs with different shapes and sizes and diverse staining shades. They also demonstrate the robustness of the algorithm in handling various morphological properties of the nucleus and cytoplasm textures.

C. Quantitative Results

We evaluate the two stages of our proposed method: the WBC detection step and the WBC segmentation step. In order to assess the detection algorithm, we estimate precision, recall and F_1-score metrics.

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

$$F_1\, measure = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{3}$$

The values for TP, FP and FN in equations (1) to (3) are derived by comparing the result of the detection algorithm with the WBC annotations. A detected WBC is counted as TP if the detection is within the annotation boundary of the corresponding WBC. Otherwise, it is counted as FP. If a WBC is not detected at all, then that misdetection will be in the FN category. In other words, the total number of annotated WBCs can be considered as the sum of WBCs correctly detected (TPs) and WBCs not recognized by the algorithm (FNs). Also, we can express the total number of detected WBCs as the sum of WBCs detected correctly (TPs) and detections made incorrectly (FPs). The aforementioned metrics are calculated for each image in the Bangladesh dataset. Each image is evaluated separately by the calculation of these metrics. The overall evaluation of our method on the entire Bangladesh dataset is computed by averaging the evaluation of all images, which results in 96.37% precision, 98.37% recall, and 97.36% F_1-score index.

To evaluate the third step of the proposed method, the segmentation step, we gauge the algorithm with the Jaccard Similarity Index and DICE Index metric.

$$Jaccard = \frac{TP}{TP + FP + FN} \tag{4}$$

$$DICE = 2 \times \frac{TP}{2TP + FP + FN} \tag{5}$$

To compute the value of TP, FP and FN in equations (5) and (6), the obtained segmentation results are compared to the manual annotations of WBCs, pixel-wise. Any segmented pixel belonging to the annotated WBC is considered a TP. If a segmented pixel falls out of the annotated WBC bounds, it is classified as FP. The pixels that belong to the annotated WBC and are not segmented are counted as FNs. For one slide image, the total number of pixels within the boundaries of the annotated WBCs is equal to the sum of pixels correctly segmented (TPs) and pixels that are not considered in the segmentation (FNs). In Fig. 9(b), the segmentation mask is overlayed on a sample image (Fig. 9(a)). The TP, FP, and FN pixels are shown as white pixels, green pixels, and pink pixels, respectively.

The overall evaluation of the segmentation step of the proposed algorithm on the entire Bangladesh dataset results in a Jaccard Similarity Index of 82.28% and a DICE Index of 78.33%.

TABLE I
QUANTITATIVE COMPARISON BETWEEN THE PROPOSED WBCS DETECTION ALGORITHM AND THE COLOR-BASED K-MEANS CLUSTERING TECHNIQUE

Metrics	K-means Clustering(%)	Proposed Algo.(%)
Precision	76.90	96.37
Recall	77.58	98.37
F1-Score	77.24	97.36

D. Comparison with a K-means Clustering Based Algorithm

Using a color-based k-means clustering algorithm for WBCs detection is an alternative approach. For example, this approach can segment colors in the image using L*a*b color space and k-means clustering. The cluster with the darkest shade can the be chosen as the one containing WBCs, or more specifically WBC nuclei. The logic behind this step is that WBCs usually appear in a darker shade of purple compared to other components in a blood smear image. Fig. 10 demonstrates this property in two examples of thin blood smear images.

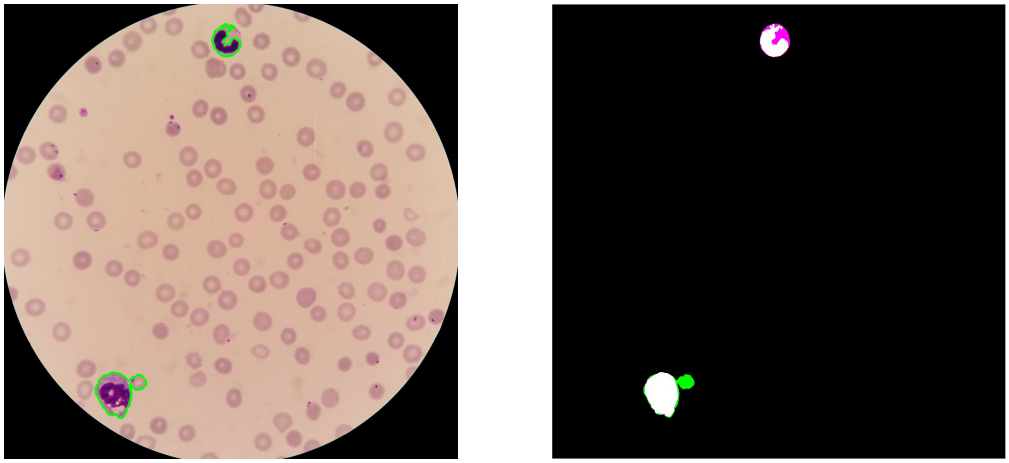
However, there can be a considerable number of cases in which WBCs are not necessarily found in the cluster with the darkest shade. Fig. 11 represents examples of blood smear images in which WBCs appear in a lighter or the same shade of purple compared with other components in the image.

This approach was tested on the entire Bangladesh dataset. The results derived from this technique are compared with the ones obtained from the proposed WBC detection algorithm in Table I.

As shown, the results of the proposed algorithm are considerably higher since the proposed approach performs robustly to the shading of the images and instead takes the contrast of pixel colors into account. In other words, the color-based k-means clustering method and the rules defined for this technique are dependent on the staining shades. Since there are different shades of staining in the Bangladesh dataset, this algorithm falls short for many cases. However, the proposed algorithm is based on the contrast rather than the color of WBCs.

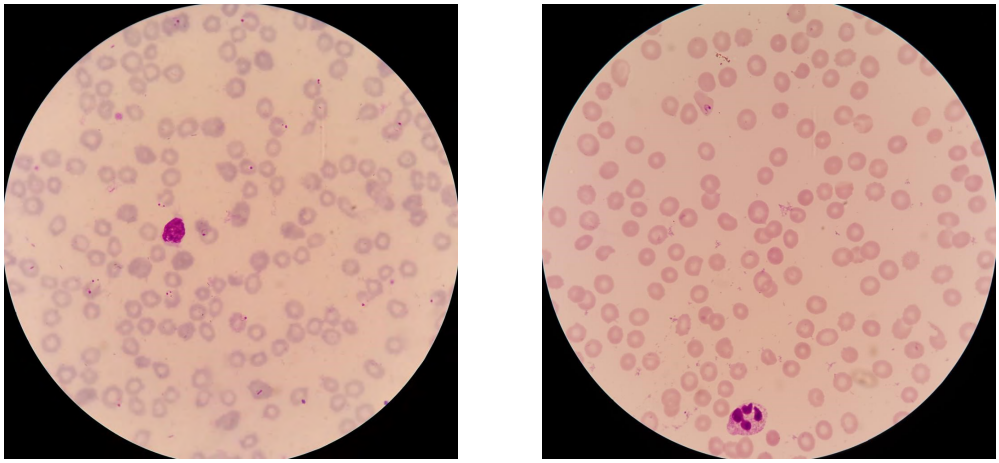
IV. CONCLUSION

We introduce an algorithm that successfully detects white blood cells in microscopic images of thin blood smears employing a range filtering approach. This algorithm can detect different types of WBCs with various shades of staining. The proposed algorithm accurately estimates the boundary pixels



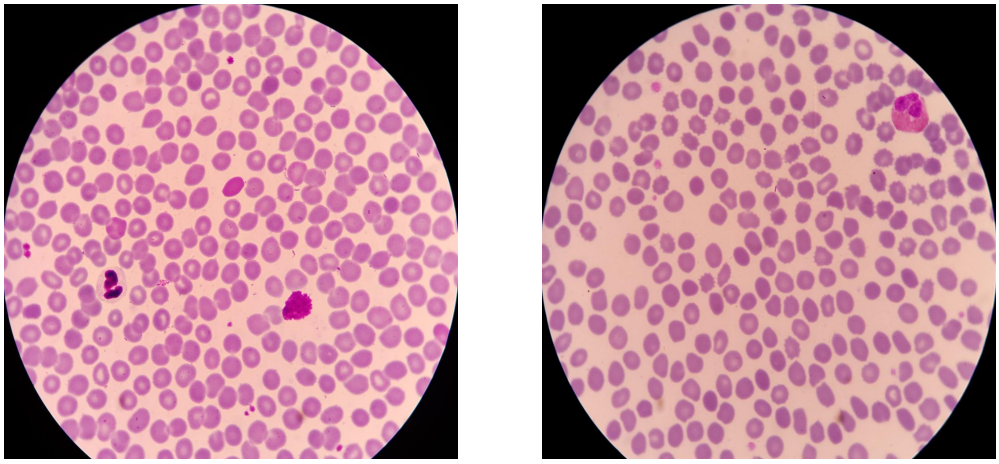
(a) (b)

Fig. 9. (a) The segmentation results marked on a sample image, (b) The segmentation mask overlayed on the sample image, with TP, FP, and FN pixel regions shown in white, green, and pink, respectively



(a) (b)

Fig. 10. Two examples of thin blood smear images and WBCs with high-contrast staining



(a) (b)

Fig. 11. Two examples of thin blood smear images and WBCs with low-contrast staining

of each cell using a level-set technique. The qualitative and quantitative results demonstrate that our algorithm can detect WBCs precisely in the first stage and can segment the detected WBCs afterwards.

V. Acknowledgment

This research is supported by the Intramural Research Program of NIH, NLM, and Lister Hill National Center for Biomedical Communications. Mahidol-Oxford Tropical Medicine Research Unit is funded by the Wellcome Trust of Great Britain.

References

[1] Seongeun Eom, Seungjun Kim, Vladimir Shin, and Byungha Ahn, "Leukocyte segmentation in blood smear images using region-based active contours," in *Advanced Concepts for Intelligent Vision Systems*. Springer, 2006, pp. 867–876.

[2] Lin Yang, Peter Meer, and David J Foran, "Unsupervised segmentation based on robust estimation and color active contour models," *IEEE Transactions on Information Technology in Biomedicine*, vol. 9, no. 3, pp. 475–486, 2005.

[3] M. Poostchi, I. Ersoy, A. Bansal, K. Palaniappan, S. Antani, S. Jaeger, and G. Thoma, "Image analysis of blood slides for automatic malaria diagnosis," 2015, p. MoPoster04.22.

[4] Carolina Reta, Leopoldo Altamirano Robles, Jesus A Gonzalez, Raquel Diaz, and Jose S Guichard, "Segmentation of bone marrow cell images for morphological classification of acute leukemia.," in *FLAIRS Conference*, 2010, pp. 86–91.

[5] Chastine Fatichah, Martin Leonard Tangel, Muhammad Rahmat Widyanto, Fangyan Dong, and Kaoru Hirota, "Interest-based ordering for fuzzy morphology on white blood cell image segmentation," *Journal of Advanced Computational Intelligence and Intelligent Informatics*, vol. 16, no. 1, pp. 76–86, 2012.

[6] Nipon Theera-Umpon, "White blood cell segmentation and classification in microscopic bone marrow images," *Fuzzy Systems and Knowledge Discovery*, pp. 485–485, 2005.

[7] Kan Jiang, Qing-Min Liao, and Sheng-Yang Dai, "A novel white blood cell segmentation scheme using scale-space filtering and watershed clustering," in *International Conference on Machine Learning and Cybernetics*. IEEE, 2003, vol. 5, pp. 2820–2825.

[8] Xiyue Hou, Mei Zhou, Zhen Sun, Qingli Li, Yingying Xue, Hongying Liu, and Yiting Wang, "A BandMax and spectral angle mapper based alogrithm for white blood cell segmentation," in *International Conference on Digital Image Processing*, 2017, vol. 10420, p. 104202B.

[9] Jaroonrut Prinyakupt and Charnchai Pluempitiwiriyawej, "Segmentation of white blood cells and comparison of cell morphology by linear and naïve Bayes classifiers," *Biomedical Engineering Online*, vol. 14, no. 1, pp. 63–82, 2015.

[10] Shubham Manik, Lalit Mohan Saini, and Nikhil Vadera, "Counting and classification of white blood cell using Artificial Neural Network (ANN)," in *International Conference on Power Electronics, Intelligent Control and Energy Systems*. IEEE, 2016, pp. 1–5.

[11] Leyza Baldo Dorini, Rodrigo Minetto, and Neucimar Jeronimo Leite, "Semi-automatic white blood cell segmentation based on multiscale analysis," *Journal of Biomedical and Health Informatics*, vol. 17, no. 1, pp. 250–256, 2013.

[12] Congcong Zhang, Xiaoyan Xiao, Xiaomei Li, Ying-Jie Chen, Wu Zhen, Jun Chang, Chengyun Zheng, and Zhi Liu, "White blood cell segmentation by color-space-based k-means clustering," *Journal of Sensors*, vol. 14, no. 9, pp. 16128–16147, 2014.

[13] Yan Li, Rui Zhu, Lei Mi, Yihui Cao, and Di Yao, "Segmentation of white blood cell from acute lymphoblastic leukemia images using dual-threshold method," *Computational and Mathematical Methods in Medicine*, vol. 2016, pp. 12, 2016.

[14] Tony F Chan and Luminita A Vese, "Active contours without edges," *Transactions on Image Processing*, vol. 10, no. 2, pp. 266–277, 2001.

[15] Zhaohui Liang, Andrew Powell, Ilker Ersoy, Mahdieh Poostchi, Kamolrat Silamut, Kannappan Palaniappan, Peng Guo, Md Amir Hossain, Sameer Antani, Richard James Maude, Jimmy Xiangji Huang, Stefan Jaeger, and George Thoma, "CNN-based image analysis for malaria diagnosis," in *International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2016, pp. 493–496.